

10/8095 fs



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

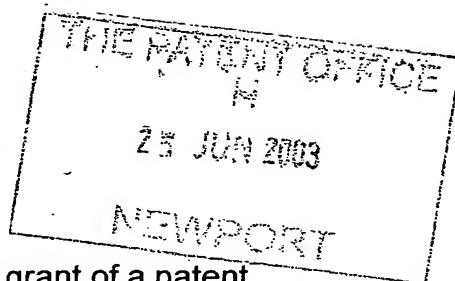
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 3 November 2003

THIS PAGE BLANK (USPTO)

Patents Act 1977
Rule 16

**The
Patent
Office**

1/77

25JUN03 E817693-1 D00611
P01/7700 0.00-0314782.4

Request for grant of a patent

The Patent Office
Concept House
Cardiff Road
Newport
South Wales NP10 8QQ

1. Your reference	GB920030013GB1		
2. Patent application number (<i>The Patent Office will fill in this part</i>)	0314782.4 25 JUN 2003		
3. Full name, address and postcode of the or of each applicant (<i>underline all surnames</i>)	INTERNATIONAL BUSINESS MACHINES CORPORATION Armonk New York 10504 United States of America		
Patents ADP number (<i>if you know it</i>)	519637001		
If the applicant is a corporate body, give the country/state of its incorporation	State of New York United States of America		
4. Title of the invention	A SEARCHABLE PERSONAL BROWSING HISTORY		
5. Name of your agent (<i>if you have one</i>)	R J BURT		
"Address for Service" in the United Kingdom to which all correspondance should be sent (<i>including the postcode</i>)	IBM United Kingdom Limited Intellectual Property Department Hursley Park Winchester Hampshire SO21 2JN		
Patents ADP number (<i>if you know it</i>)	0919006		
6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (<i>if you know it</i>) the or each application number	Country	Priority App No (<i>if you know it</i>)	Date of filing (<i>day/month/year</i>)
7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	No of earlier application	Date of filing (<i>day/month/year</i>)	

8. a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:
 a) any applicant named in part 3 is not an inventor, or
 b) there is an inventor who is not named as an applicant,
 or
 c) any named applicant is a corporate body.)

Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description	10
Claim(s)	2
Abstract	1
Drawing(s)	5

JWL

10. If you are also filing any of the following, state how many against each item.

Priority documents -

Translations of priority documents -

Statement of inventorship and right to grant of a patent (Patents Form 7/77) 2

Request for preliminary examination and search (Patents Form 9/77) -

Request for substantive examination (Patents Form 10/77) -

Any other documents (please specify) -

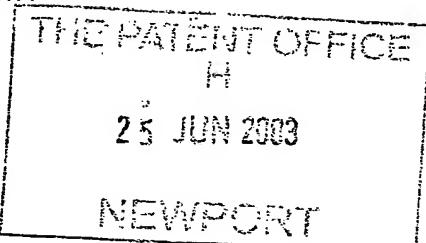
11. I/We request the grant of a patent on the basis of this application

Signature

23 June 2003
Date

12. Name and daytime telephone number of person to contact in the United Kingdom

T ROBERTSON
01962 817424



The
Patent
Office

7/77

Patents Act 1977
Rule 15

Statement of inventorship and of right to grant of a patent

The Patent Office
Concept House
Cardiff Road
Newport
South Wales NP10 8QQ

1. Your reference	GB920030013GB1
2. Patent application number (if you know it)	0314782.4 25 JUN 2003
3. Full name of the or of each applicant	INTERNATIONAL BUSINESS MACHINES CORPORATION
4. Title of invention	A SEARCHABLE PERSONAL BROWSING HISTORY
5. State how the applicant(s) derived the right from the inventor(s) to be granted a patent	By virtue of employment and agreement
6. How many, if any, additional Patents Forms 7/77 are attached to this form?	
7.	I/We believe that the person(s) named over the page (and on any extra copies of this form) is/are the inventor(s) of the invention which the above patent application relates to.


Signature

24 June 2003
Date

8. Name and daytime telephone number of
person to contact in the United Kingdom

T Robertson
Tel: 01962 817424

Enter the names, addresses and postcodes of the inventors in the boxes and underline the surnames

Arjan DE MES
Resident of The Netherlands
c/o IBM United Kingdom Limited
Intellectual Property Law
Hursley Park
Winchester
Hampshire SO21 2JN
England

Patents ADP number (*if known*)
8560482001

If there are more than three inventors, please write their names and addresses on the back of another Patents Form 7/77 and attach it to this form

REMINDER

Have you signed the form?

Patents ADP number (*if known*)

A SEARCHABLE PERSONAL BROWSING HISTORY

Field of the Invention

5 The invention relates to the field of Internet technologies and in particular to creating a searchable personal browsing history whilst accessing over a communications network a plurality of data resources.

Background of the Invention

10 The World Wide Web (WWW) has evolved into a very useful tool for searching for information, banking on-line, shopping on-line, booking a holiday and checking share prices. The WWW comprises millions of individual webpages and it soon becomes easy to loose track of which web pages have been
15 visited when trying to locate a particular piece of information.

An example of this is searching the WWW using a search engine such as Google (Google is a registered trademark of Google Technology Inc) or Yahoo (Yahoo is a registered trademark of Yahoo! Inc.), for a topic such as
20 knowledge management. The search results are displayed as a list of titles and hyperlinks to knowledge management websites.

If a particular hyperlink is selected from the search results a web page is displayed. Embedded within this web page may be a variety of other
25 hyperlinks which direct a user to further knowledge management web pages which may or may not be of interest to the user. Once the user has found the web page with the information that they need the user can either print, download or bookmark the web page for future reference.

30 The above method of saving, printing or downloading, works well for the information located by the user at that particular moment in time. A problem occurs when days, weeks or months later the user is triggered into remembering a piece of information that they read whilst navigating their way through the numerous websites and did not save, print or download a
35 particular web page containing that information.

Unless the user actively bookmarks every web page that they visit the user is unable to remember what they read or where a particular web page can be found. Typically users rely on a search engine to re-find a web page that they read. This becomes a complicated and tedious task and may not work if a key intermediate web page has been amended or deleted.

40 A common approach to saving web pages for later use is to use a cache. Most web browsers maintain a cache of recently visited web pages and other web

resources in the client device's local file system, using an HTTP request to check with the original server that the cached web pages are the most current pages available before displaying them in the web browser. A web browser cache suffers from the disadvantage that it is of an uncontrolled and temporary nature that requires periodic scanning/indexing in order for the information stored in the cache to be of any use to a user. Further, some documents are never placed in the cache and therefore it does not give a full indication of the web pages or web resources that a user has accessed over a particular period of time.

10

Another method of storing recently visited web pages is to save the web pages for off-line viewing. This facility is offered in current versions of Microsoft Internet Explorer (Microsoft Internet Explorer is a registered trade mark of Microsoft Corporation in the U.S and other countries). To save a visited web page for off line viewing a user can bookmark the web page currently being accessed. Microsofts Internet Explorer provides a 'wizard' presenting the user with a number of options to customise the content for off line viewing.

20

A disadvantage with the above approach is that a user has to actively select the web pages to be bookmarked and be aware that the web page will be needed in the future.

25

Another approach can be found in a paper written by Manber U et al (to appear in 1997 Usenix Technical Conference..., Jan 6-10, 1997), (web reference <http://webglimpse.org/pubs/webglimpse/pdf>) from the Department of Computer Science, University of Arizona, Tucson. The paper discusses a tool called WebGlimpse which analyses collections of webpages. WebGlimpse analyses a given WWW archive for example a website, a collection of specific documents or a private history cache and computes neighbourhoods i.e. the most relevant documents according to a user's specification. Once this has been completed search boxes are added to selected pages, remote pages are collected if relevant and the pages are cached locally. Users are able to browse the website using any of the added search boxes. A disadvantage of this approach is that a user has to actively indicate to WebGlimpse that the user wishes to archive a particular website or a particular web page. If a user is suddenly triggered into remembering something that they read days or weeks ago and the web page has not been archived, the user still must try and retrace their steps using their preferred search engine.

30

35

40

Yet another approach is discussed in a paper entitled 'Lifestreams: organising your electronic life' written by Freeman, E et al, from the department of Computer Science, Yale University, New Haven, United States.

5 Lifestreams describes a system which provides a time ordered stream of documents which functions as a diary of a persons electronic life. The paper describes creating a time ordered stream of documents starting with a person's electronic birth certificate and the time-ordered document stream moving towards the present day with more current document that the user has added to the time-ordered document stream.

10 A disadvantage of using the approach offered by Lifestreams is that a user must actively create a document which is subsequently added to the time-ordered document stream. This approach is not suitable for saving web pages for off-line viewing because the user is required to actively indicate which web pages are to be saved.

15 Therefore an improved method and system is required for storing a plurality of web resources accessed over a network by a user, and for displaying to the user, the accessed web resources in a meaningful way.

Disclosure of the Invention

20 Accordingly, in a first aspect, the present invention provides a method for creating a searchable personal browsing history, the method comprising the steps of: requesting a data resource from a device in a communications network; extracting metadata and textual data from the received data resource; indexing the extracted metadata and textual data and storing the indexed metadata and textual data in a data store; and displaying a searchable personal browsing history.

25 An advantage of using the above approach is that each data resource that a user accesses is isolated such that the metadata and textual data can be extracted and stored in a data store. There is no active input required by the user i.e. the user does not have to actively select that a data resource should be saved. Thus, the present invention provides an accurate account of the data resources accessed over a communications network by the user. The user may define the types of categories to be displayed in the searchable personal browsing history thereby personalising the data displayed. Further, a user may search the searchable personal browsing history and thereby creating a view within the searchable personal browsing history defined by the search results and one or more user defined categories.

40 In one embodiment extracted metadata and textual data are stored with a reference to the data resource's original location. Thus, preventing a need for a complete copy of the data resource to be stored in a data store.

5 In one preferred embodiment a calculation is performed on the extracted metadata to create statistical information relating to a user's browsing activity when accessing across a network the data resource. An advantage of this approach is that a user is able to view their browsing activity in categorised views which provides efficient access to the required information.

10 Preferably the calculated statistical information provides a user with categories of recently visited web pages, most frequently visited web pages, recently visited downloads and/ or recently visited images.

15 According to a second aspect, the present invention provides a system for creating a searchable personal browsing history, the system comprising: a proxy component for inspecting over a communication network a requested data resource; a search/index component for extracting and indexing metadata and textual data from a received data resource; and a presentation component for displaying the browsing history.

20 The proxy component inspects all of a users network traffic and is not selective in the network traffic that the proxy component inspects. Each data resource is isolated and is passed to an index/search component to allow the metadata data and textual data to be extracted from the data resource. The system described above, allows the process to be automatic with no input required from the user.

25 Brief description of the drawings

30 The invention will now be described by way of example only, with reference to the accompanying drawings, in which:

35 Figure 1, illustrates the searchable personal browsing history method running on a data processing system, according to a preferred embodiment of the present invention;

40 Figure 2, illustrates the components of the personal browsing history according to a preferred embodiment of the present invention;

45 Figure 3, depicts according to a preferred embodiment of the present invention, a flow chart illustrating the operational steps carried out by a system when browsing, over a communication network one or more data resources;

Figure 4, depicts according to a preferred embodiment of the present invention, a flow chart illustrating the operational steps performed when creating a searchable personal browsing history; and

5 Figure 5, depicts a user's searchable personal browsing history according to a preferred embodiment of the present invention.

Detailed description of the preferred embodiments of the Invention

10 Figure 1 is a block diagram of a data processing environment in which the preferred embodiment of the present invention may be advantageously applied. In figure 1, a client/server data processing host 100 is connected to other client/server data processing host 135 and 140 via a network 130, such as, for example, the Internet. Client/server data processing host 100 has a processor 105 for executing programs that control the operation of the client/server data processing host 100, a RAM volatile memory element 110, a non-volatile memory 120, and a network connector 115, for use in interfacing with the network 130 for communication with the other client/servers 135 and 140.

20 The personal browsing history application 125 may be deployed on the client/server data processing host 100 as a standalone client application interfacing with a user's browser and accessing over a network 130 data resources requested from client/servers data processing hosts 135 and 140.

25 Alternatively, the personal history application may be deployed as a server application on client/server data processing hosts 135 or 140 allowing client/server data processing host 100 to access the personal history application 125 over the communication network 130. For the remainder of this document the personal browsing history application 125 will be described as being deployed as a client application on the client/server data processing host 100 and accessing over a communication network 130, a plurality of data resources requested from client/server data processing hosts (herein referred to as a web server) 135 and 140.

30 35 Figure 2 illustrates the components that make up the personal browsing history application 125; such components include a proxy component 200, a search/index 205 component and a presentation component 210. Each of these components will be discussed in turn.

40 The proxy component 200 allows the personal browsing history application 125 to keep a local representation of recently accessed data resources. These data resources may be web pages, graphics, downloads or any other resource that are accessed over the network 130.

The proxy component 200 determines, on receipt of a request for a data resource, whether it can handle the request itself or if another proxy server must be contacted to additionally handle the request for the data resource.

5

The latter situation can occur in a corporate environment where requests for data resources outside of the corporate Intranet are configured to be sent to a proxy server before allowing access to the Internet. If the proxy component 200 determines that it can handle the request for a data resource directly, the proxy component 200 accesses the network 130 and contacts the web server 140 to serve the data resource. The web server 140 sends the request back to the proxy component 200 residing on the host 100. Once the request is received by the proxy component 200 the request is sent to the user's browser and the index/search component 205 begins to process the data resource.

10

The storing of a representation of an accessed data resource requires no active input from the user, it is carried out automatically by the index/search component 205 when the proxy component 200 inspects each accessed data resource.

15

The personal browsing history application 125 further comprises an index/search component 205 which extracts metadata and textual data from a data resource and indexes the extracted data to form a textual index for searching.

20

To enable data to be displayed through a conventional browser a mark up language such as HTML is used to specify the formatting, presentation and the text and images that comprise the contents of a web page. A typical piece of HTML tagging is as follows

```
30 <html>
31   <head>
32     <meta name="keywords" content="corporate home page" />
33   <title>My Company</title>
34   </head>
35   <body TEXT="000000" BGCOLOR="FFFFFF" leftmargin=0 topmargin=0 marginwidth=0
        marginheight=0> The body tag specifies how to display the text and graphics
        to a user.
40   <h1>This is a heading tag </h1>
41   <p>The start of a new paragraph</p>
42   </body>
43 </html>
```

When the index/search component 205 receives a data resource from the proxy component 200, the index/search component traverses each of the html tags and extracts metadata and textual data from the data resource. Examples of metadata are: the URL of the web page, the last modified date, fields specified as metadata in the HTML, the title of the web page, and the amount of text on the web page specified in a word count. The textual data i.e. the natural language information embedded in the web page between a body tag (<body></body>) is extracted. Both metadata and textual data are stored with a reference to the original location of the data resource. The reference to the original location of the data resource may comprise an HTTP request or other appropriate protocol.

The personal browsing history application 125 further comprises a presentation component 210 for displaying a searchable personal browsing history created by the personal history application 125.

Referring to Figure 3 it will now be explained how the personal browsing history application 125 operates when accessing a network 130 such as the Internet.

At step 300 the user accesses the network using the personal browsing history application 125 configured to work with the user's browser. A web page or other web resource such as a downloadable file or graphic image may be accessed in the normal manner by entering in a Uniform Resource Locator (URL) into the URL address input box in the user's browser. The browser sends a request message to the proxy component 200 and the proxy component 200 determines whether it can handle the request itself or whether another proxy must handle the request. If the proxy component 200 can handle the request itself, a request for a data resource is sent through the network 130 to the web server 140 or 135 depending on which web server can serve the requested data resource specified by the URL.

The web server 135 or 140 looks up the path name of the requested data resource and sends back the data resource in a reply message through the network 130 to the personal browsing history application 125. At step 320 the proxy component 200 forwards the requested resource to the web browser, where it is loaded into the browser window and displayed to the user at step 325.

At step 305 the index/search component 210 extracts metadata and textual data from the contents of the data resource as described previously. The metadata and the textual data extracted by the index/search component 210 are used to dynamically create a searchable personal browsing history which displays the user's browsing activity when accessing over a network 130

data resources. The metadata and the textual data extracted in step 305 are stored in a data store at step 310.

At step 315 the stored metadata and textual data are indexed to reflect any recently stored metadata and textual data in step 310. A reference to the data resource's original location is also stored at step 310 such that the extracted metadata and the textual data create a textual index along with a reference to the data resource's original location. Each time the proxy component 200 receives a requested resource the textual index is updated to reflect the addition of a new data resource.

The stored metadata and textual data are indexed each time a data resource is accessed over the network 130 thereby allowing the user to constantly view and search the data resources that they have accessed.

Step 320 is carried out in parallel with steps 305, 310, and 315. The requested data resource is returned to the browser and displayed to the user at step 325. The above steps allows the personal history browsing application 125 to work in the background constantly extracting, storing and re-indexing the extracted metadata and textual data, whilst the user is browsing the WWW.

Considering now how the personal browsing history may be used, a user may be triggered into remembering something that they read some time ago. The user remembered they read it but have no idea when or where. Referring to Figure 4, a user locates a data resource that the user had previously accessed by loading the presentation component 210 from a menu option within the user's browser. The user's browser sends a request to the proxy component 200 and the proxy component 200 loads the presentation component into the user's browser to display the searchable personal browsing history.

At step 400 the proxy component 200 loads the user settings for the searchable personal browsing history. The user settings define information about how the user would prefer the searchable personal browsing history to be personalised. The user settings are defined in a user profile and may be modified at any time by the user. The user settings consist of information such as for example, which sections may be displayed in the presentation component 210, granting access rights to the personal history application 125 to other user's and password settings. Usability settings may include for example, the colour of the text to be displayed in the presentation component within the user's browser when viewing the searchable personal browsing history.

5 The metadata and textual data that was extracted from the accessed data resource at step 305 of Figure 3 are retrieved from the data store. The metadata is used to calculate statistical information on the activity of a user accessing over a network 130 a plurality of data resources. The type of calculations that may be performed enable the determination of the most recently visited web pages at step 410, the most frequently visited web pages at step 415, the most recently downloaded files by the user at step 420, and the most recently downloaded images by the user at step 425. Thus, the statistical information allows a user to see their past browsing 10 activity categorised by the type of calculation performed.

15 At step 405 the user is able to perform a key word search in the index of the stored metadata and textual data. The keyword search is performed by typing search criteria into a search input box. The index/ search component 205 uses the search criteria to locate and retrieve the information requested by the user. At step 430 the personal browsing history application 125 creates a searchable personal browsing history which is tailored to the search results, the statistical information and the configuration settings as defined by the user and displayed at step 435. 20 The searchable browsing history may contain the results of multiple searches (iterations of step 405) and their results.

Step 435 will now be explained further with reference to Figure 5.

25 Figure 5 illustrates a searchable personal browsing history as generated by the personal browsing history application 125. The searchable personal browsing history is a dynamic view changing each time the user performs a new search on the index in step 405 of Figure 4 or accesses over a network 130 one or more data resources.

30 The searchable personal browsing history comprises several different sections, recently visited sites 500, favourite sites 510, downloaded files 515, image downloads 520 and a search sections 525 and 530 for inputting a search criteria.

35 In the search section 525 the example search criteria shown is '+ "web services" -.net'. The searchable personal browsing history locates within the indexed data, all references to "web services" and scores the results according to the most relevant.

40 The scoring is displayed to the user by a colour gradient bar 505, the higher the score the more intense the colour. The scoring is defined by the metadata extracted from the web resource at step 305 of Figure 3. The search results in each section depend on the information contained within

the metadata and in the textual data thereby, displaying information that is only relevant to the user's browsing activity.

5 The user is therefore able to dynamically see which web resources they have visited at a particular point in time and quickly locate the information they had seen before. The searchable personal browsing history dynamically updates the view every time the user visits another web page or downloads a file or image.

CLAIMS

1. A method for creating a searchable personal browsing history, the method comprising the steps of:

5 extracting metadata and textual data from a received data resource in a network;

10 indexing the extracted metadata and textual data and storing the indexed metadata and textual data in a data store; and

15 displaying the searchable personal browsing history.

2. A method as claimed in claim 1 wherein the extracted metadata and textual data are stored with a reference to the original location of the data resource.

20 3. A method as claimed in claim 1 wherein the searchable personal browsing history is updated each time metadata and textual data is extracted from a received data resource.

4. A method as claimed in claim 1 wherein the indexed metadata and textual data form a textual index for searching for the data resource.

25 5. A method as claimed in claim 1 wherein a calculation is performed on the extracted metadata to create statistical information relating to a user's browsing activity when accessing across the network the data resource.

30 6. A method as claimed in claim 5 wherein the statistical information comprises recently visited web pages, most frequently visited web pages, recently visited downloads and recently visited images.

35 7. A system for creating a searchable personal browsing history, the system comprising:

a proxy component for inspecting over a communication network a requested data resource;

40 a search/index component for extracting and indexing metadata and textual data from a received data resource; and

a presentation component for displaying the searchable personal browsing history.

8. A computer program product directly loadable into the internal memory of a digital computer, comprising software code portions for performing the steps of any one of claim 1 to claim 6 when said product is run on a computer.

9. A web hosting service for providing a searchable personal browsing history, the web hosting service comprising:

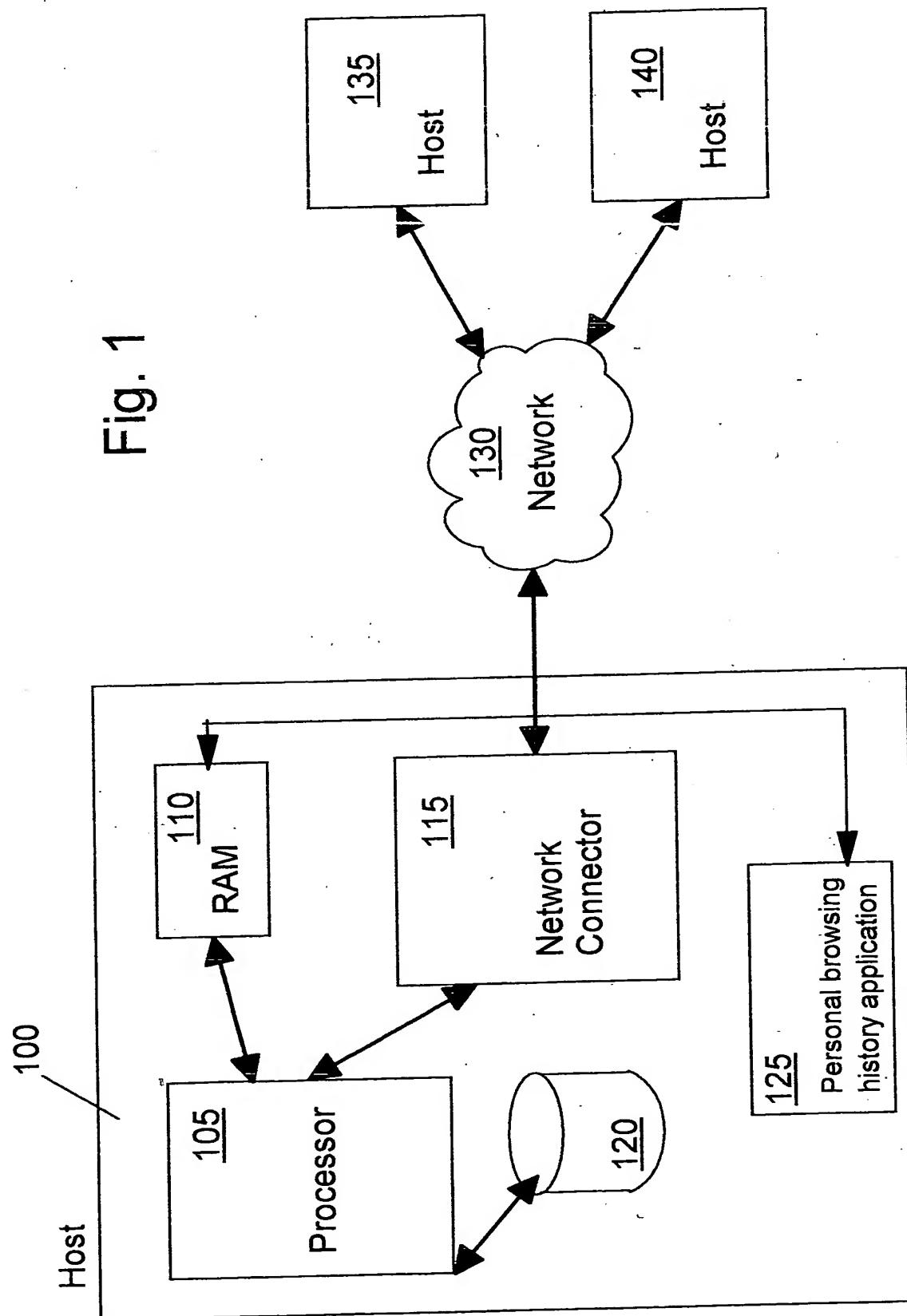
providing a data resource from a device in a communications network;
extracting metadata and textual data from the data resource;
indexing the extracted metadata and textual data and storing the indexed metadata and textual data in a data store; and
displaying the searchable personal browsing history.

ABSTRACT**A SEARCHABLE PERSONAL BROWSING HISTORY**

5 A method for creating a searchable personal browsing history, comprising
requesting a data resource from a device in a communications network;
extracting metadata and textual data from the received data resource;
indexing the extracted metadata and textual data and updating the indexes
10 stored in a data store; and displaying a searchable personal browsing
history categorised by a user defined criteria.

THIS PAGE BLANK (USPTO)

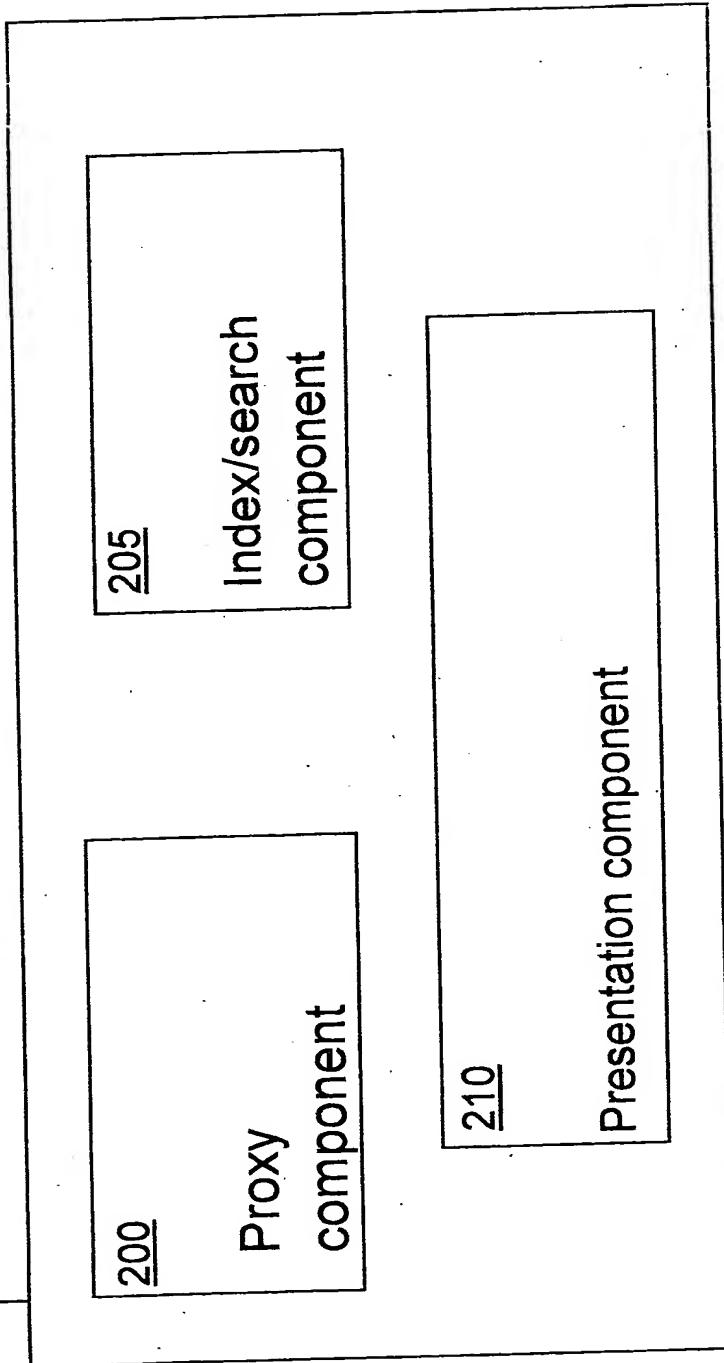
Fig. 1



THIS PAGE BLANK (USPTO)

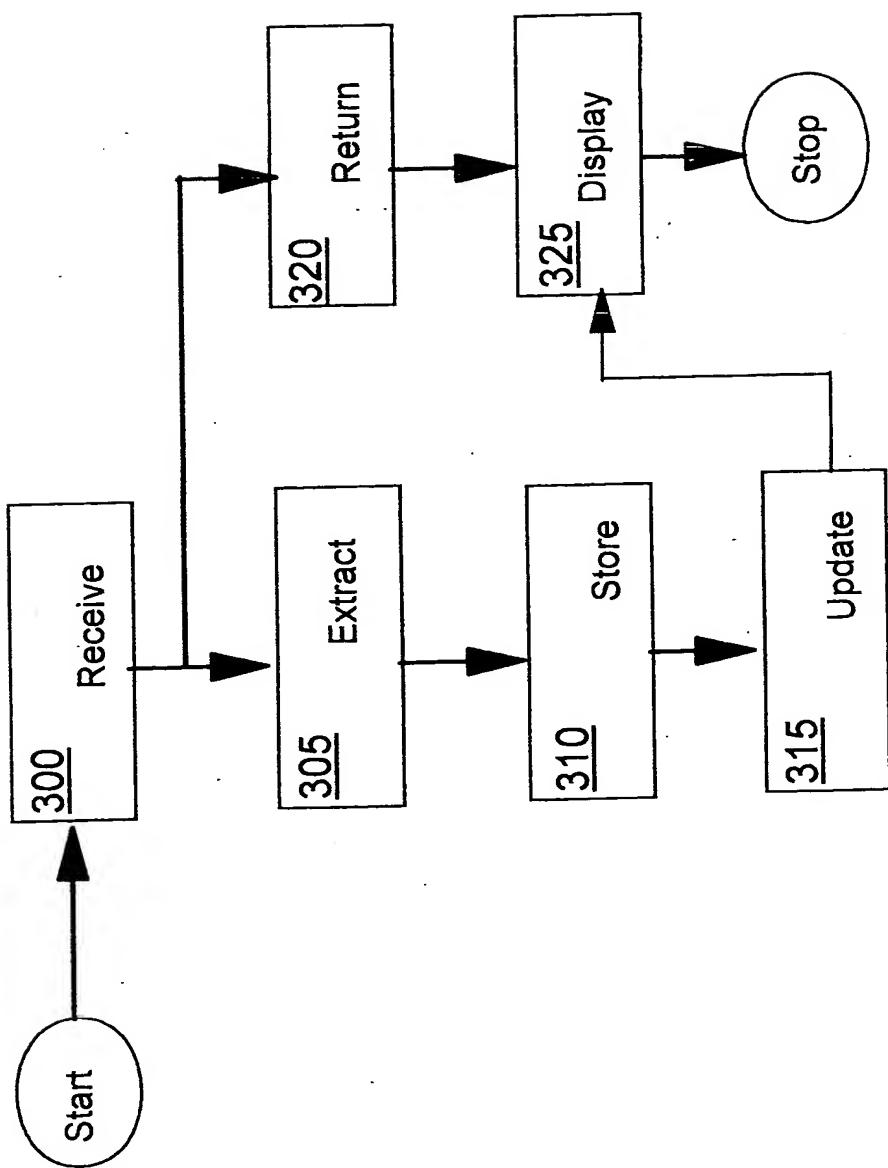
125 Personal browsing history application

Fig. 2



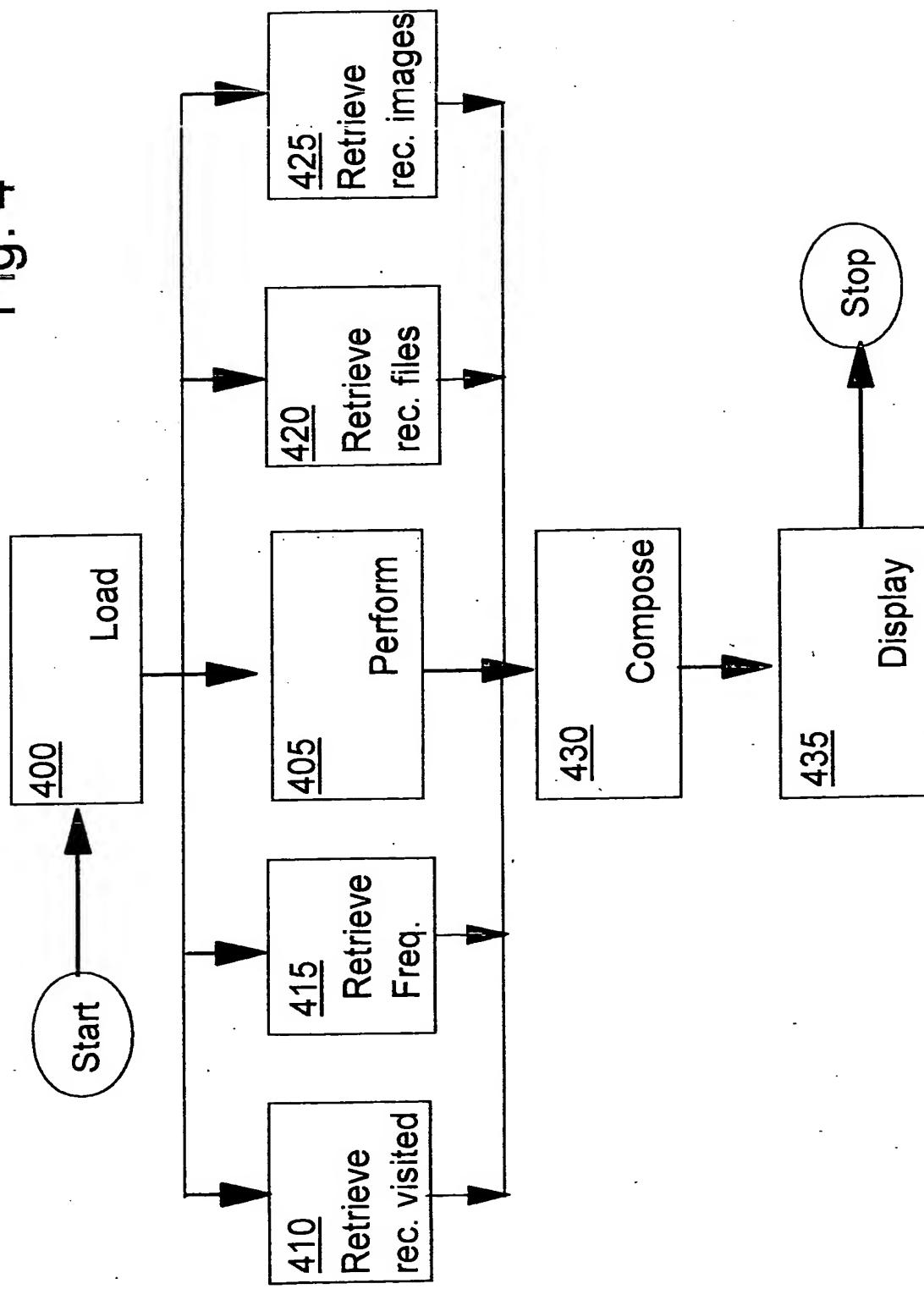
THIS PAGE BLANK (USPTO)

Fig. 3



THIS PAGE BLANK (USPTO)

Fig. 4



THIS PAGE BLANK (USPTO)

Fig. 5

515

Personal Homepage

RECENT SITES

This is the lastest list of sites recently visited. Color indicates time since last visit.

500

505

510

FILE DOWNLOADS

These are the recently downloaded files.

- tspace133.zip
- ie5setup.exe
- suse7-bootimage.tar.gz
- netscape7a.exe
- mozilla.zip

more ▶

IMAGE DOWNLOADS

These are the recently downloaded images.

- dont1.jpg (w3.ibm.com)
- 20020520_lbm_top_gated_fet.jpg (resw5.research.ibm.com)
- logo2.gif (www.nytimes.com)
- hh.gif (www.tweakers.net)
- masthead3.gif (ing.cmpnet.com)
- title.gif (images.slashdot.org)
- topmozilla.gif (images.slashdot.org)

more ▶

FAVORITE SITES

These are the sites that are most frequently visited. Color indicates frequency.

520

525

530

WEB SERVICES

+ "Web Services" - net

These are the search results. Color indicates score against the search criteria.

525

530

Configure

Google

Slashdot: News for nerds, stuff that matters

Informatic, website voor de informatieverziening

Nederlands Genootschap voor Informatica

IBM Intranet

IBM Research

more ▶

SEARCH

New SEARCH

Search for keywords in documents you visited. Currently, historical data is kept for 365 days.

5/5

THIS PAGE BLANK (USPTO)